

METHOD AND SYSTEM FOR AUGMENTING WEB CONTENT

10

INVENTORS
Douglas Stevenson
Craig Allen Gooding

15

PRIORITY CLAIM

This application claims priority to provisional patent application 60/484,818 filed on

July 2, 2003.

BACKGROUND OF THE INVENTION

Hypertext is the organization of computer-based text into connected associations enabling a user to quickly access information that the user chooses. An instance of such an association is called a hyperlink or hypertext link. Hypertext was the main concept that led to the invention of the World Wide Web, which is nothing more than an enormous amount of information content connected by an enormous number of hyperlinks.

25

While the hyperlink has proven to be a successful means of relating two pieces of information, the process of generating hyperlinks has proven to be generally tedious. To create a single link, the process requires an author to define such portions, usually a text string or photo element of a structured file (usually a text document, web page, or other form of document) from which the hyperlink originates and a destination address at which the hyperlink terminates. In a closed system such as in a local network of workstations, the destination might be within the same file, directory, or computer, or the destination may be a designated file within a designated directory on the network.

30

Even with the required information, the knitting together of hyperlinks still requires some skill. The would-be author of a document with suitable links must first identify the



25315

PATENT TRADEMARK OFFICE

XTEM-1-1013AP

816 Second Avenue
Seattle, Washington 98104
206.381.3300 • F: 206.381.3301

content of the file the author seeks to augment and then must use an appropriate application to edit the file. Generally, the MIME header embedded in the file identifies the file type. The embedded header allows a computer software product to recognize the data by virtue of its Multi-Purpose Internet Mail Extensions (“MIME”) type. MIME is an extension of the 5 original Internet e-mail protocol that lets people use the protocol to exchange different kinds of data files on the Internet including audio, video, images, application programs, and other kinds, including text generally in the ASCII format. Once identified, the file is opened for review using the appropriate application as identified by the MIME header.

When the network is broader, such as on the Internet, static addresses on the Internet 10 may be used as destinations. A Uniform Resource Locator is the address of a file (resource) accessible on the Internet. The type of resource depends on the Internet application protocol. Using the World Wide Web’s protocol, Hypertext Transfer Protocol (HTTP), the resource can be an HTML page, an image file, a program such as a common gateway interface 15 application or Java applet, or any other file type supported by HTTP. The URL contains the name of the protocol required to access the resource, a domain name that identifies a specific computer on the Internet, and presents a hierarchical description of a file location on the computer.

For this reason, the content of resources may not have all of the links that would be 20 useful. Old data might direct browsers to addresses where no data is now stored. “Link rot,” describes a gradual loss of data at URL’s linked to documents. This gradual loss occurs when a destination document is removed while the link in an originating document to the destination document remains. The reader receives a “404 message,” an arbitrarily assigned 25 code indicating that the page to which the reader has directed the browser no longer exists at the designated address. Another form of link rot occurs when the destination page has been changed in content and is no longer relevant according to the sending description.

Where parties, such as advertisers, wish to inject links into existing resources in order 30 to direct the reader’s browser to designated resources, it is critical that the links remain current. Because files are static, links that are old will not complete the hyperlink transit and therefore will lose the benefit of the hyperlink. Fixedly embedding links in the file subjects the file to link rot.

There is, therefore, an unmet need in the art for a publishing system and method for augmenting resources and maintaining suitable current hyperlinks within the resources.

SUMMARY OF THE INVENTION

A system for augmenting data from a source data file with data from a reference 35 database to generate an augmented data file is provided. The system includes a reference



25315

PATENT TRADEMARK OFFICE

XTEX-1-1013AP

- 2 -

BLACK LOWE & GRAHAM PLLC

816 Second Avenue
Seattle, Washington 98104
206.381.3300 • F: 206.381.3301

database including at least one reference datum. A handler component is configured to retrieve a source data file including the structured datum. A locator component is configured to locate the structured datum in the source data file; an analyzer component configured to associate the identified structured datum to one reference datum to create an association 5 according to an analyzing strategy. A generating component is configured to generate a hyperlink based upon the association and embeds the generated hyperlink in the source file to create an augmented data file.

One presently preferred embodiment includes a system for augmenting file content, including web content, with hyperlinks to designated destinations. The system works based 10 on finding a datum (a data subset of a file) within a file, recognizing the datum based upon the contents of a reference database, associating the datum with a designated resource (in a presently preferred embodiment by means of a uniform resource locator address) and generating a hyperlink in the data source file.

The generated hyperlink (in a presently preferred embodiment) receives a user-friendly name based on the contents of the resource located at the uniform resource locator 15 address. The an embodiment is a add-on to a browser allowing the browser to augment files “on the fly,” i.e. where the user directs the browser to a resource located on a network, the method analyzes the file as it is opened by the browser, augments the file with appropriate hyperlinks, and displays the augmented file with active hyperlinks. “Clicking on” the 20 hyperlink will redirect the browser to the associated uniform resource locator address.

Another presently preferred embodiment provides a rigorous procedure for augmenting files that assures that a greater number of hyperlinks are more uniformly applied than is the rule with human authoring of hyperlinks. A reference database can check the content of the file and will always place a hyperlink where appropriate based upon context. 25 Because the reference database can be readily updated, the invention assures that redirection to current resources, preventing the dead link, i.e. “Error 404, file not found.”

A presently preferred embodiment provides a method and a software product to add advertisements to existing web content by hyperlinking occurrences of structured data such as text strings to resources located at a uniform resource locator address.

As will be readily appreciated from the foregoing summary, the invention provides a 30 system and a method for rapidly and for rigorously augmenting files with hyperlinks.

BRIEF DESCRIPTION OF THE DRAWINGS

The preferred and alternative embodiments of the present invention are described in detail below with reference to the following drawings.

35 FIGURE 1 is a flowchart of the method for augmenting source data files;



25315

PATENT TRADEMARK OFFICE

XTEN-I-1013AP

- 3 -

BLACK LOWE & GRAHAM PLLC

816 Second Avenue
Seattle, Washington 98104
206.381.3300 • F: 206.381.3301

FIGURE 2 is a block diagram of a system for augmenting source data files;
FIGURE 3 is screen shot of a source data file selected for augmentation;
FIGURE 4 is a block of text selected for augmentation;
FIGURE 5 is the block of text selected for augmentation based upon the occurrence
5 of a text string;
FIGURE 6 is an excerpt of an exemplary reference database;
FIGURE 7 is a screen shot of the source data file selected for augmentation showing
hyperlink and associated user-friendly name;
FIGURE 8 is a screen shot of the contents of a destination file upon redirection of the
10 browser.

DETAILED DESCRIPTION OF THE INVENTION

By way of overview, a system for augmenting data from a source data file with data
from a reference database to generate an augmented data file is provided. The system
includes a reference database including at least one reference datum. A handler component
15 is configured to retrieve a source data file including the structured datum. A locator
component is configured to locate the structured datum in the source data file; an analyzer
component configured to associate the identified structured datum to one reference datum to
create an association according to an analyzing strategy. A generating component is
configured to generate a hyperlink based upon the association and embeds the generated
20 hyperlink in the source file to create an augmented data file.

FIGURE 1 is a flowchart of the method 9 for augmenting source data files. At an
appropriate starting terminus 10, the method begins by reading a structured datum from a
source data file at a block 13. The source data file may be one designated by an input
uniform resource locator address or by any suitable means to designate a resource. Upon
25 opening, the method 9 may optionally identify the type of content on the page with a content
identifier such as a MIME header. In one embodiment of the invention, the method 9 merely
searches for the presence of a reference datum, either informed by the content identifier or by
simply searching an occurrence of a well-structured datum within a given file. However,
once the file is open, the method has the contents of the file available for comparison to a
30 reference database.

At a block 16, the method 9 locates an occurrence of a reference datum corresponding
to the structured datum read in the source data file. One presently preferred means of
discerning the correspondent relation between the structured datum and the reference datum
is by a JavaScript call made to a web-enabled database. The Java script then compares the
35 contents of the source data file with reference data stored in a web-enabled reference



25315
PATENT TRADEMARK OFFICE

XTEN-1-1013AP

database. In one presently preferred embodiment, the reference and structured data are keywords. The JavaScript code then extracts the text from the document and converts all of the keywords in the document to hyperlinks.

Other analyzing strategies are also available. Though keywords are a facile and efficient means of generating hyperlinks. One presently preferred embodiment uses a “fuzzy expert” or a neural network analysis of the source data file, such as by a natural language search of the document to generate a distinct identifier for the content in the source data file. One advantage of a natural language search is the ability to better place content in context making links more contextually appropriate, for instance, security might relate to security of a physical plant such as security of a residence in one source data file in one context and security of a website in another. Natural language searches, however, create a large processing overhead, making them less desirable where such resources are at a premium.

At a block 19, the method 9 generates an association based upon the located reference datum in the reference database. The reference datum will include not only the identifier, such as a keyword in one embodiment, but also the associated uniform resource locator address of the intended destination based upon the occurrence of the identifier in the source data file. Generating an association means to hyperlink the located structured datum in the source data file to the associated uniform resource locator as found in the reference datum in the reference database. The generated hyperlink might optionally include a user-friendly description of the content of the resource found at the associated uniform resource locator address and additionally might include an additional identification code such as an “advertiser id.” In each embodiment, the generated hyperlink is added to the original source data file at a block 22 resulting in an augmented data file. Having generated the augmented data file, the method 9 then terminates at a block 25.

FIGURE 2 is a block diagram of a system for augmenting source data files. In this exemplary system, a network 33 is shown in a presently preferred embodiment. Those skilled in the relevant art will readily appreciate that the system may be practiced without the presence of a network link. Also a reference database 39 is shown as directly connected to the locator 42 and the analyzer 45. The system is not compromised by network links to any of the several components. One presently preferred embodiment has the reference database 39 set on a web-enabled page for remote calls through the Internet to the database. Like the presence of the network 33 described above, the absence of a defined link through the Internet does not compromise the operation of the method.

A source data file 30 resides on a server on a network 33. A handler 36 retrieves the source data file 30 for use by the system. A locator 42 examines the retrieved source data



25315

PATENT TRADEMARK OFFICE

XTEM-1-1013AP

- 5 -

BLACK LOWE & GRAHAM PLLC

816 Second Avenue
Seattle, Washington 98104
206.381.3300 • F: 206.381.3301

file 30 for comparison to the reference database 39 according to an analyzing strategy. The locator 42 designates found structured data from the source data file 30 and found reference data from the reference database 39 and provides the reference data to an analyzer 45.

The analyzer 45 is used to create associations between each found structured datum and the uniform resource locator address within the corresponding reference datum found by the locator 42. These associations define the nature of a hyperlink a generator 48 generates according to the association created at the analyzer 45. The generator 48 embeds these hyperlinks in the source data file 30. The resulting augmented data file 50 is returned to the handler 36 to reside at a uniform resource locator address on the network 33.

FIGURE 3 is a screen shot 101 of contents of a source data file 30 (FIGURE 2) selected for augmentation. For the purposes of this exemplary discussion, a block of text 104 is selected by the method for augmentation. The present invention does not require the opening of the source data file 30 in a browser however for the purposes of illustration, the screen shot 101 is provided herein.

FIGURE 4 is a block of text 107 selected for augmentation. The locator 42 (FIGURE 2) begins its analysis of the selected block by isolating the block of text 107 as structured data.

FIGURE 5 is the block of text 107 selected for augmentation indicating the occurrence of a text string 111. The locator 42 isolates the occurrence of the text string 111 because of its presence in the reference database 39.

Referring to FIGURES 2 and 6, while the excerpt 115 of the exemplary reference database 39 shows it to be a “flat file” database, any database 39, relational, flat file, or other configuration will suitably fulfill the basic functions of associating an identifier such as a text string 118, with a uniform resource locator address 124, and optionally a user-friendly description 121 of the contents of the file found at the associated uniform resource locator address 124. (Not shown is the optional “advertiser id.”)

The locator 42 (FIGURE 2) refers to the database as it reviews the contents of the source data file 30 (FIGURE 2). According to the analyzing strategy, the text string 118 occurs at a reference datum in the source data file 30, the locator 42 provides the uniform resource locator 124 associated with the found text string 118 along with the occurrence of the text string in the reference datum to the analyzer 45. Upon receiving the ordered pair from the locator address 42, the analyzer 45 creates the association. With the association from the analyzer 45, the generator 48 creates a hyperlink. As a result, the text block receives the appropriate hyperlink.



25315

PATENT TRADEMARK OFFICE

FIGURE 7 is a screen shot 101 of the augmented data file 50 showing the selected block of text 104, hyperlink 129, and associated user-friendly name 133. As will readily be appreciated by those skilled in the relevant art, the augmented data file 50 need not reside on a network. For instance, where a browser is enhanced with the inventive method 9 (FIGURE 1), the browser might be directed to content on the Internet. Upon opening the source data file 30 (FIGURE 2), the browser will effect the method 9 (FIGURE 1) such that rather than displaying the contents of the source data file 30, the browser will display the augmented data file 50. The resulting augmented files 50 would be current at the time of viewing. In a publisher embodiment, the advertiser might continually files on the network, 10 replacing augmented files on the network.

The augmented data file 50 is displayed as set forth in FIGURE 7. When a reader places the mouse cursor over the hyperlink 129, a new layer for display showing the associated user-friendly name 133 may be optionally evoked. The “advertiser id” may be optionally embedded but not necessarily visible, the advertiser id. This advertiser id 15 specifically provides a means for tracking the number of times the hyperlink 129 is activated and generating a record for tracking revenue due to advertising. The reader clicks on the hyperlink 129 to direct the browser to a destination file.

FIGURE 8 includes a screen shot 140 of the browser displaying the contents of the destination file located at the uniform resource locator address 137 upon redirection of the browser. When the reader clicks on the hyperlink 129, a new browser window is opened and 20 directed through a click-tracking server to the hyperlink destination.

While the preferred embodiment of the invention has been illustrated and described, as noted above, many changes can be made without departing from the spirit and scope of the invention. For example, the method 9 will generate static files on the network at distinct 25 uniform resource locator addresses 124 and 137 in order to distinguish from the original source data files. Accordingly, the scope of the invention is not limited by the disclosure of the preferred embodiment. Instead, the invention should be determined entirely by reference to the claims that follow.



25315

PATENT TRADEMARK OFFICE